Deception in Democracy: Political Lying Accusations and Their Effects on Democratic Citizenship (DEMO-LIES)

Postdoctoral Scholarship Holder in Computational Political Science

Dr. Bastián González-Bustamante

✉ b.a.gonzalez.bustamante@fgga.leidenuniv.nl

Presentation at the DEMO-LIES Search Committee
November 10, 2025

Work Package 2 Proposal

When Political Blame Becomes Accusations of Lying

Blame Shifting in Presidential Systems: Ministerial Terminations' Corrective Effect on Approval

Bastián González-Bustamante*
Leiden University
Universidad Diego Portales

August 16, 2025

This article has been accepted for publication in *Public Opinion Quarterly* published by Oxford University Press

Abstract

How do ministerial terminations affect presidential approval? Presidents face unexpected challenges related to stochastic events such as scandals, policy failures or economic crises. We argue that the termination of ministers who have received calls for their resignation presents an opportunity for the president to send signals to the electorate in the expectation of a corrective effect on popularity through a blame-shifting dynamic. The central argument is that this dynamic only occurs in coalition governments where political responsibility may be more easily attributed to the coalition's different parties and factions, weakening personalisation centred on the president and facilitating blame shifting and the corrective effect. The expectation of a corrective effect on approval is tested using instrumental variables (IV) regressions applied to novel data on ministerial terminations and resignation calls in 124 governments in 12 presidential democracies. The data were gathered by combining data mining, machine learning techniques and survey marginal time series based on the dyad ratios algorithm for approval. The main findings support

*Postdoctoral Researcher in Computational Social Science and Lecturer in Governance and Development, Institute of Public Administration, Faculty of Governance and Global Affairs, Leiden University, Netherlands. Wijnhaven, Turfmarkt 99, The Hague 2511 DP, Netherlands, b.a.gonzalez.bustamante@fgga.leidenuniv.nl, https://bgonzalezbustamante.com, ORCID iD https://orcid.org/0000-0003-1510-6820. Lecturer in Public Administration, Faculty of Administration and Economics, Universidad Diego Portales, Chile. bastian.gonzalez.b@mail.udp.cl.

1

This (**preliminary**) proposal focuses on the strategic deployment of liar accusations in political discourse, connecting WP1 data to party-level incentives and contexts.

Core question. When do parties deploy liar accusations, and who targets whom?

Liar accusations could be a sharpened form of **blame shifting** that parties use under specific conditions.

Expectations

Blame Shifting/Avoidance

- Parties under pressure divert attention through credibility attacks, shifting focus from policy failures to opponent trustworthiness.

Responsibility Cues

- Responsibility cues make governments natural targets, as incumbents “own” policy outcomes and face accountability pressures.

Electoral Timing

- Accusations intensify as elections near, particularly from opposition parties seeking to undermine incumbent credibility.

Key Hypotheses: Who Accuses Whom?

- **Incumbents' Hypothesis.** Incumbents are more frequently accused, especially in pre-election periods, because they bear responsibility for outcomes.
- **Opposition Hypothesis.** Opposition parties are the main senders of accusations in pre-election periods, using credibility attacks as strategic weapons to challenge government legitimacy.
- **Populist Hypothesis.** Populist parties are likely to accuse because parliamentary norms less deter them.
- **Ideological Hypothesis.** Ideologically extreme parties are more likely to accuse, particularly when targeting ideologically distant opponents.

Data Structure and Measurement

WP1 data are aggregated to party-year and directed party-party-month (or quarter) for dyadic analysis.

- **Sender volume.** How frequently each party (member) makes accusations.
- **Receiver volume.** How frequently each party (member) is accused.
- **Dyadic counts.** Party-to-party accusation patterns.

Covariates include government/opposition status, ideology and extremity, populism scores, vote share, months-to-election, issue topics, and country-year controls.

Artwork by DALL·E 3 model

Empirical Strategy

Artwork by DALL·E 3 model

Multilevel NBRMs

Party random intercepts with country and year FEs to account for nested data structure.

Overdispersion Handling

NBRMs are preferred over Poisson due to expected overdispersion from unobserved heterogeneity (IJPP, 2024).

Robustness Checks

Pseudo-election dates as placebos, recodifications, and uncertainty propagation from WP1 classification tasks.

Empirical Strategy

Why not classic linear mixed models?

On counts is problematic, but we can also try to use proportions.

Why not zero-inflated?

Only in cases in which there is a structural “zero” barrier (e.g., parties legally impeded from making accusations).

Potential problems

How to control retaliation? Use lags and interactions for moderation analyses carefully.

Artwork by DALL·E 3 model

Expected Takeaways

Artwork by DALL·E 3 model

Expected patterns

- **9 months before.** Parties begin escalating accusations.
- **6 months before.** Sharp climb in opposition accusations. Incumbent receiver rates rise.
- **Election period.** Peak accusation intensity, especially from populist and extreme parties.

Extensions

- Linking this to fact-checking databases as a check on accusation validity and strategic deployment.
- Discontinuity and specific stochastic events.

Finding the Needle in a Haystack

General Overview

Finding the Needle in a Haystack:
Using Large Language Models to Detect Rare Speech Acts

Michal Mochtak*[1] and Maurits J. Meijers[2]

[1]Department of Political Science, Radboud University, Nijmegen, Netherlands,
https://orcid.org/0000-0001-5598-5642,
https://scholar.google.com/citations?user=Lq_sKXEAAAAJ
[2]Department of Political Science, University of Antwerp, Antwerp, Belgium,
https://orcid.org/0000-0001-8034-1910,
https://scholar.google.com/citations?user=D4-vnXIAAAAJ

October 24, 2025

Abstract

This paper presents a cost-effective and scalable approach for identifying rare speech acts, using political disinformation accusations as a case study. We propose a pipeline designed to process large text corpora efficiently while optimizing the use of limited human annotation resources. Rare and varied in form, from overt claims like "fake news" to subtler expressions such as "bending the truth", disinformation accusations pose a significant challenge for manual detection, especially in formal settings like parliamentary discourse. Our pipeline addresses this by combining automated techniques with manual annotation to uncover and analyze these elusive acts. The results demonstrate the pipeline's effectiveness in capturing a wide range of disinformation-related speech, offering a promising solution for studying infrequent yet important communicative phenomena in political language.

Keywords: large language models; supervised machine learning; automated text analysis; parliamentary discourse; rare speech act detection

*Corresponding author: Michal Mochtak, Nijmegen School of Management, Department of Political Science, Radboud University, PO Box 9102, 6500 HC Nijmegen, Netherlands; email: michal.mochtak@ru.nl

Disclaimer. Some comments might be implemented, but others are rather future work.

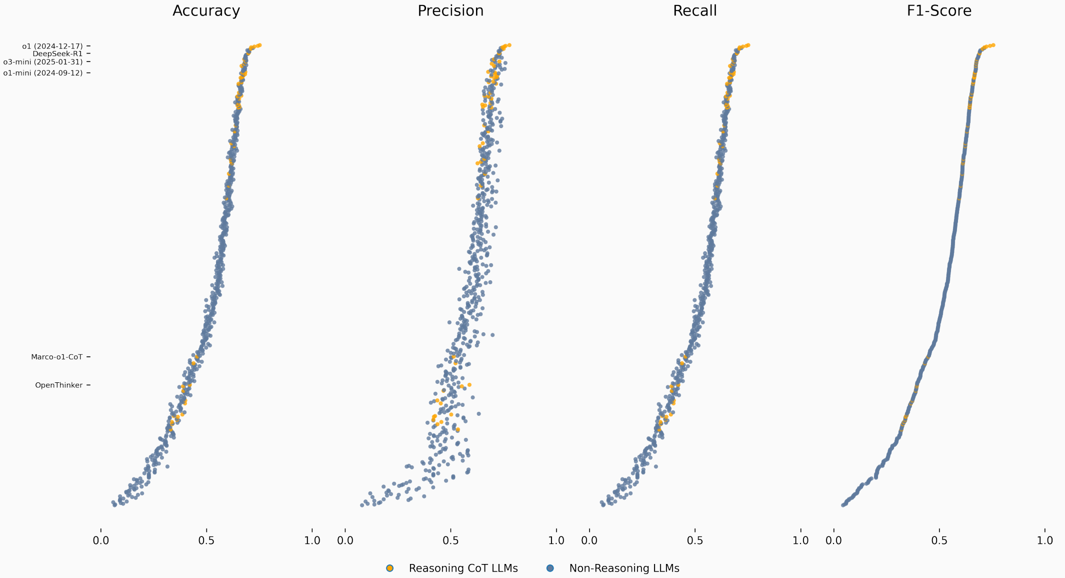
Strong points. Robust methodology that combines human-in-the-loop (HITL) and advanced techniques. Strong showcase of computational social science application. **LieLine is a significant contribution**, and the argument of imbalanced data is appealing.

Room to improve. The manual annotation and validation section could benefit from more precise explanations to enhance reader comprehension.

Specific Comments

- RoBERTa is okay, however, you may want to try **ModernBERT**. I have fine-tuned some models for CAP coding, and ModernBERT is performing really well.
- I understand that you did not try XLM-RoBERTa because you use English as an anchor with MT. I guess that was because of the data annotation and validation steps in the pipeline.
- Using English as an anchor is not a problem at all, but it does reduce the linguistic scope of the work slightly. A natural next step would be to **analyse multilingual texts**.
- I would suggest **reporting MT metrics** (i.e., BLEU or similar).
- Using **GPT-3.5-Turbo is a bit problematic** because (a) it is a closed model and (b) currently outdated. You may want to try (for future work) SOTA models beyond a few-shot classification. Although this is not a big issue because you put HITL just after this in the pipeline.

LLMs for Multilingual Policy Agenda Topic Annotation



Comparing with Fine-Tuned Models

Language	Best LLM	F1-Score	Fine-Tuned	F1-Score	Δ Val	Δ Best LLM
Danish	GPT-4.5	0.679	Babel Machine	0.925	+0.065	+0.246
Dutch	o1	0.724	Babel Machine	0.906	+0.066	+0.182
English	o1	0.706	Babel Machine	0.869	−0.031	+0.163
French	o1	0.714	Babel Machine	0.821	−0.029	+0.107
Hungarian	GPT.4-5	0.672	Babel Machine	0.751	−0.099	+0.079
Italian	o1	0.675	Babel Machine	0.930	+0.120	+0.255
Portuguese	o1	0.651	Babel Machine	0.867	−0.063	+0.216
Spanish	o4-mini	0.756	Babel Machine	0.916	+0.066	+0.160

Note. All estimates are weighted F1-scores obtained on our fixed held-out test set. The columns Δ Val and Δ Best LLM indicate: (i) the change relative to the best result on the model's own validation set; and (ii) the change relative to the strongest zero-shot LLM, respectively.

How much data leakage is present here? Probably something, so that the results may be inflated. However, the potential advantages of fine-tuning still seem to outweigh in-context learning, even for BERT-like models. * Babel Machine by Sebők et al. (2024).

Fine-Tuned BERTs

Fine-tuned XLM-RoBERTa
F1 validation 0.819
F1 held-out set 0.810
https://doi.org/10.57967/hf/6863

Fine-tuned ModernBERT
F1 validation 0.831
F1 held-out set 0.809
https://doi.org/10.57967/hf/6864

More Specific Comments

- You may want to mention the **debate about open, closed LLMs and reproducibility** issues. This is a controversial point, however, from my viewpoint, it is not significant because current LLMs, although they do not offer full reproducibility, do exhibit a slight variation during an iteration (less than 5%). But proprietary models may not be welcomed by many reviewers.
- You only have **inter-coder measures for the pilot** ($n = 100$, $\alpha = 0.72$). You compensate to some extent with your cross-validation experiments. The bootstrap is particularly interesting. However, the lack of a gold standard in this part may be a loose end for the peer-review process.
- The downstream application with **LieLine is a strong point**, and the events potentially associated with discontinuities could open new research (i.e., migration crisis, Brexit, rule of law problems). I would only suggest reporting in the Hugging Face model the **carbon footprint** of the training (and inference) if it is possible, and also some **average inference times** with the model using CPU vs GPU.

Thank you very much!